

Modélisation Et Prévision Mensuelle De La Salinité Par Réseaux De Neurones De Type Perceptron Dans La Partie Centrale Du Canal Du Mozambique

Audiat Miller HERINDRAIN¹, Jean Eugène RANDRIANANTENAINA², RATSIMAVO Jacques Chrysologue³, FATIANY Pierre Ruphin⁴, LALAOHARISOA Sahoby⁵, Adolphe A. RATIARISON⁶

^{1,2,3,4}Ecole Doctorale Géosciences, Physique, Chimie de l'Environnement et Systeme Hotes-Patogenes (GPCEHP) de l'Université Tuléar Madagascar

^{5,6}Laboratoire Dynamique de l'Atmosphère, du Climat et des Océans (DYACO), de l'Université d'Antananarivo Madagascar

¹millerhery1@gmail.com, ²rjendria02@yahoo.fr, ²rat.chryso@gmail.com, ⁴fpierrezuphin@yahoo.fr

⁵sahoby81@gmail.com, ⁶adolphe.ratiarison@univ-antananarivo.mg

Auteur correspondant : Audiat Miller HERINDRAIN. E-mail : millerhery1@gmail.com



Résumé : Cette étude présente une méthodologie hybride combinant régression linéaire multiple et Perceptron Multicouche (MLP) pour la prévision mensuelle de la salinité de surface dans le canal du Mozambique. Appliquée aux données de la région (1980-2018), l'approche identifie les précipitations comme forçage dominant du bilan hydrique régional. Le modèle neuronal final, optimisé selon une architecture à 4 couches cachées, présente une corrélation de $R=0,67$ et une RMSE de 0,060 PSU avec les observations, démontrant sa précision quantitative. Les projections pour 2019-2028 suggèrent une tendance à l'augmentation de la salinité dans le canal. Cette méthode offre un cadre robuste pour la surveillance et la prévision de la variabilité hydro-climatique de cette voie maritime stratégique.

Mots-clés : Salinité de surface, Canal du Mozambique, Modèle hybride, Réseau de neurones, Régression linéaire, Prévision océanique.

Abstract: This study presents a hybrid methodology combining multiple linear regression and Multilayer Perceptron (MLP) for monthly forecasting of surface salinity in the Mozambique Channel. Applied to data from the region (1980-2018), the approach identifies precipitation as the dominant forcing factor in the regional water balance. The final neural model, optimized according to a 4-layer hidden architecture, shows a correlation of $R=0.67$ and an RMSE of 0.060 PSU with observations, demonstrating its quantitative accuracy. Projections for 2019-2028 suggest a trend toward increasing salinity in the channel. This method provides a robust framework for monitoring and forecasting hydro-climatic variability in this strategic sea lane.

Keywords: Surface salinity, Mozambique Channel, Hybrid model, Neural network, Linear regression, Ocean forecasting.

Introduction

La salinité de surface de l'océan joue un rôle fondamental dans la circulation thermohaline, les échanges air-mer et la variabilité climatique. Sa variabilité spatiale et temporelle influence de manière significative les écosystèmes marins, le climat régional et global, ainsi que les activités humaines liées au milieu marin. Dans ce contexte, la prévision fiable de la salinité représente un enjeu majeur pour la compréhension et la gestion des environnements océaniques.

Dans le canal de Mozambique, région océanographiquement dynamique, la salinité présente une forte variabilité saisonnière et interannuelle. Les approches basées sur les réseaux de neurones artificiels offrent une alternative performante pour la modélisation de relations non linéaires complexes.

I. Délimitation de la zone d'étude

La zone d'étude couvre la partie centrale du canal de Mozambique, délimité par 35°30'E à 43°30'E et 23°9' S et 19°9' S (Fig1). Cette région correspond au secteur suffisamment éloigné des côtes africaines et malgaches pour limiter l'influence directe des processus côtiers.

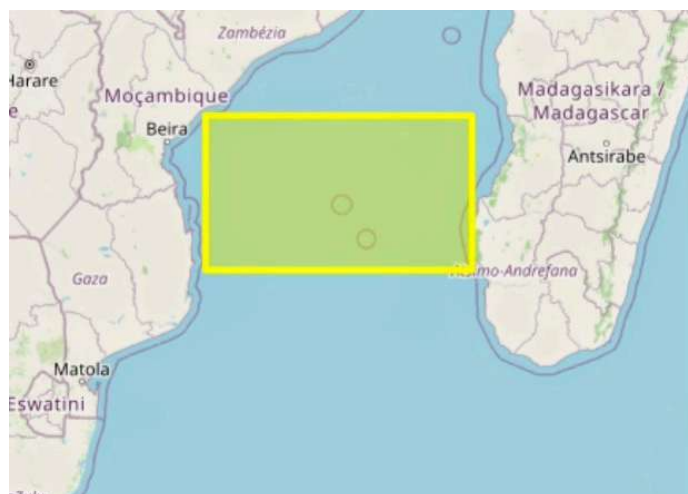


Figure 1 : Zone d'étude (GeoJson)

II. Présentation des données utilisée

Cette étude s'appuie sur cinq variables climatiques et océaniques essentielles représentées sous forme de séries chronologiques continues sur une grille géographique régulière (1980-2018). Les variables analysées et leurs unités respectives sont les suivantes :

- **Température de Surface de la Mer (SST)** : degrés Celsius (°C) ;
- **Pression au niveau de la mer (SLP)** : hectopascals (hPa) ;
- **Précipitations (P)** : millimètres par jour (mm. jour⁻¹) ;
- **Salinité de Surface (SSS)** : unité pratique de salinité (PSU) ;
- **Évaporation (E)** : millimètres par jour (mm. jour⁻¹) .

III. Prétraitement des données

Avant toute analyse et modélisation, un prétraitement systématique a été appliqué à l'ensemble des jeux de données journalières (SST, pression, Précipitations, Salinité, Évaporation). L'étape principale a consisté à calculer des moyennes mensuelles à partir de ces données quotidiennes, afin d'obtenir des séries temporelles homogènes pour l'analyse et la modélisation.

IV. Méthodologie de Modélisation et de Prévision

La prévision de la salinité de surface est abordée via une architecture de modélisation hybride innovante. Cette approche en deux étapes combine un préfiltrage linéaire par régression multiple avec un modèle non linéaire de type Perceptron Multicouche (MLP). L'objectif est de tirer parti de la simplicité et de l'interprétabilité d'un modèle linéaire pour extraire les relations dominantes, tout en

utilisant la puissance d'approximation universelle du MLP (Hornik et al., 1989) pour capturer les interactions complexes et non linéaires résiduelles entre les forçages climatiques.

✓ Étape 1 : Préfiltrage et pondération linéaire

Une régression linéaire multiple est d'abord calibrée pour établir une relation de base entre les forçages climatiques (prédicteurs) et la salinité observée (variable cible) :

$$SSS(t) = \beta_0 + \beta_P P(t) + \beta_E E(t) + \beta_{SLP} SLP(t) + \beta_{SST} SST(t) + \epsilon(t)$$

où t indexe le temps (mois), β_i sont les coefficients de régression partielle estimés par la méthode des moindres carrés ordinaires, et $\epsilon(t)$ représente les erreurs résiduelles. Les coefficients standardisés β_i (obtenus en régressant sur les variables normalisées) quantifient l'importance relative de chaque prédicteur dans la variance expliquée linéairement de la SSS.

Ces coefficients sont ensuite utilisés pour créer un jeu de données d'entrée pondéré et physiquement informé pour le modèle neuronal :

$$\mathbf{X}(t) = [\beta_P \cdot P(t), \beta_E \cdot E(t), \beta_{SLP} \cdot SLP(t), \beta_{SST} \cdot SST(t)]^T$$

Cette transformation a pour effet d'amplifier le signal des forçages les plus pertinents et d'atténuer celui des forçages peu corrélés, servant de guide physique initial au modèle non linéaire (Bishop, 2006).

✓ Étape 2 : Architecture et apprentissage du Perceptron Multicouche (MLP)

La modélisation non linéaire proprement dite est effectuée par un réseau de neurones artificiels de type Perceptron Multicouche (MLP). L'architecture générale d'un tel réseau, représentée de manière schématique dans la Figure 2, repose sur une structure *feedforward* à couches multiples et totalement connectées, qui constitue un cadre éprouvé pour l'approximation de fonctions complexes (Cybenko, 1989).

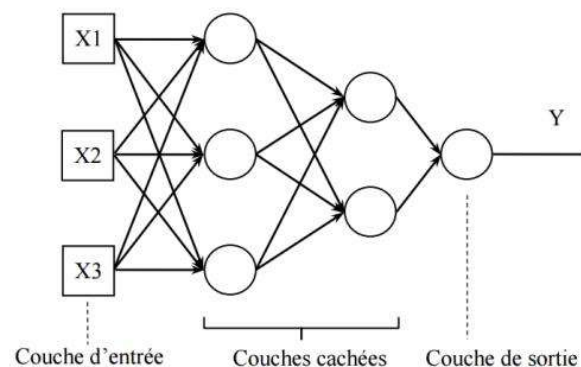


Figure 2 : Le modèle de perceptrons multicouches (MLP)

Conformément à la **Figure 2**, notre MLP est organisé en trois blocs principaux :

- La **couche d'entrée**, représentée par les neurones X_1 à X_4 , correspond aux quatre composantes du vecteur d'entrée pondéré $\mathbf{X}(t)$. Chaque neurone transmet simplement la valeur de la variable d'entrée correspondante.
- Une ou plusieurs **couches cachées** (une seule est représentée dans le schéma pour plus de clarté), comprenant un nombre n de neurones. Chaque neurone de cette couche reçoit l'intégralité des sorties de la couche précédente. Il calcule

d'abord une combinaison linéaire de ses entrées, à laquelle s'ajoute un terme de biais, avant d'appliquer une fonction d'activation non linéaire $\phi(\cdot)$:

$$h_j(l)(t) = \phi\left(\sum_i w_{ji}(l)a_i(l-1)(t) + b_j(l)\right)$$

où $h_j(l)$ est l'activation du $j^{\text{ème}}$ neurone de la couche l , $a_i(l-1)$ est la sortie du $i^{\text{ème}}$ neurone de la couche précédente, $w_{ji}(l)$ sont les poids synaptiques, et $b_j(l)$ les biais. Dans notre implémentation, nous utilisons la fonction d'activation ReLU (Rectified Linear Unit), $\phi(z)=\max(0,z)$, pour sa capacité à atténuer le problème de gradient évanescent et à accélérer la convergence (Glorot et al., 2011).

- **La couche de sortie**, composée d'un unique neurone Y , génère la prédiction finale de salinité $SSS(t)$. Pour ce problème de régression, une fonction d'activation linéaire est utilisée sur ce neurone final :

$$\overline{SSS}(t) = \sum_j w_j^{(out)} h_j^{(L)}(t) + b^{(out)}$$

où L désigne la dernière couche cachée.

V. Résultats

V.1 Équation de la régression linéaire multiple

L'analyse par régression linéaire multiple a permis d'établir une relation quantitative entre la salinité de surface (variable dépendante) et les quatre forçages climatiques indépendants. L'équation obtenue, avec tous les coefficients significatifs au seuil de confiance de 95% ($p < 0.05$), est la suivante :

$$S = 35.11276 - 0.35459P + 0.04053E - 0.00113R + 0.00339T$$

Où :

- S : Salinité de surface [PSU]
- P : Pluviométrie (précipitations) [mm jour⁻¹]
- E : Évaporation [mm jour⁻¹]
- R : Pression au niveau de la mer [hPa]
- T : Température de surface de la mer (SST) [°C]

La constante du modèle ($K = 35.11276$ PSU) représente la salinité moyenne de référence lorsque tous les forçages sont nuls, et intègre implicitement l'influence d'autres facteurs non inclus dans le modèle.

Cette équation peut être réécrite sous forme d'une équation de prédiction des anomalies de salinité (ΔS) en soustrayant la constante et en considérant les variables comme des anomalies par rapport à leurs moyennes climatologiques respectives :

$$\Delta S = -0.35459P + 0.04053E - 0.00113R + 0.00339T$$

Où ΔS représente la variation de salinité prédite [PSU] à partir des quatre paramètres climatiques. Chaque terme correspond à une variable d'entrée pondérée par son coefficient de régression partielle :

- $-0.35459P$: Contribution des précipitations (effet négatif dominant)
- $+0.04053E$: Contribution de l'évaporation (effet positif modéré)

- $-0.00113R$: Contribution de la pression (effet négatif faible)
- $+0.00339T$: Contribution de la température (effet positif faible)

V.2 Architecture neuronale optimale

Après un processus d'apprentissage et de validation, l'architecture de réseau de neurones retenue est un Perceptron Multicouche (MLP) dont la **couche d'entrée** reçoit les **quatre variables climatiques préalablement pondérées** par leurs coefficients de régression linéaire (P^* , E^* , SLP^* , SST^*). Le modèle comporte **quatre couches cachées** (15, 13, 20 et 20 neurones). Cette configuration, illustrée dans la **Figure 3**, offre le meilleur équilibre entre précision et capacité de généralisation, surpassant les autres architectures testées.

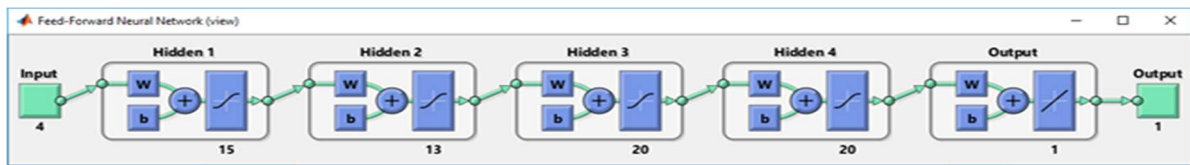


Figure 3 : Architecture neuronale de la salinité de l'océan mensuelle dans la zone d'étude

V.3 Validation du modèle

La validation du modèle neuronal montre un coefficient de corrélation de **$R = 0,67$** entre les valeurs observées et prédites, confirmant une relation significative. L'erreur quadratique moyenne (RMSE) de **$0,0600$ PSU** atteste quant à elle de la précision quantitative des prédictions. Cette combinaison d'une corrélation modérée et d'une faible erreur valide la performance globale du modèle.

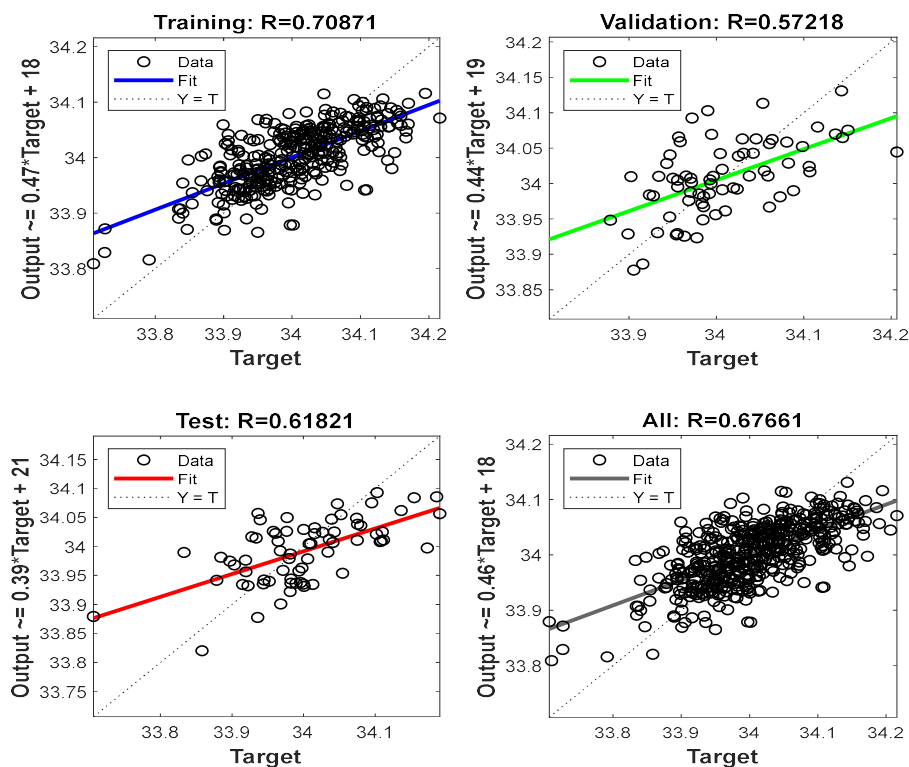


Figure 4 : Corrélation entre sortie calculée et sortie désirée de la salinité de l'océan dans la zone d'étude

V.4 Résultats visuels et projections à moyen terme

La **Figure 3** présente une synthèse des performances du modèle hybride, confrontant les observations historiques aux projections futures. Cette étape valide la capacité du modèle à passer d'une phase d'apprentissage à une phase de prédiction opérationnelle.

A partir de ce modèle, on peut prédire la variation de la salinité mensuelle dans cette zone pour les 10 années futures (2019-2027). La Figure 4 illustre les données d'observations (noir), l'apprentissage du réseau (bleu) et la prédiction de cette salinité (mauve).

Modélisation de salinité mensuelle dans la zone de etude (1m à 15m de profondei

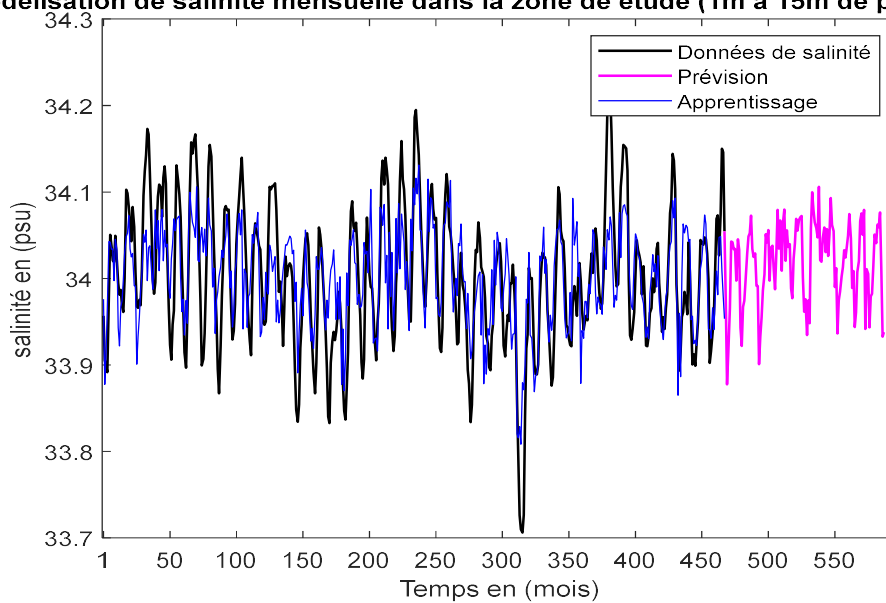


Figure 5 : Modélisation de la salinité de l'océan da la zone d'étude

Le tableau ci - dessous résume les prévisions annuelles pour 2019-2028.

Tableau de prévision pour la salinité

	Janv	Fév	Mar	Avr	Mai	Juin	Juil	Aoû	Sep	Oct	Nov	Déc
2019	33,98	33,88	33,91	33,97	34,04	34,04	34,03	34,03	34,00	34,05	34,03	33,94
2020	33,92	33,97	33,97	34,00	34,03	34,05	34,05	34,07	34,03	34,03	34,03	33,99
2021	33,99	33,90	33,93	33,95	33,99	34,00	34,05	34,04	34,06	34,00	34,04	34,04
2022	34,01	33,99	34,03	34,08	34,03	34,07	34,01	34,07	34,08	34,02	34,06	34,04
2023	34,04	33,99	33,98	34,02	34,04	34,07	34,06	34,07	34,07	34,07	34,00	33,96
2024	33,98	33,93	33,97	33,95	34,06	34,10	34,08	34,06	34,06	34,05	34,11	34,02
2025	34,02	34,05	34,06	34,03	34,06	34,03	34,05	34,09	34,06	34,06	33,99	34,01
2026	33,94	33,97	33,98	34,00	34,01	34,03	34,02	34,04	34,05	34,07	34,05	33,99
2027	34,01	33,94	33,96	34,01	34,02	34,07	34,08	34,05	34,08	33,94	34,01	34,01
2028	33,94	33,98	33,98	34,01	34,05	34,04	34,06	34,06	34,08	34,01	33,93	33,94

Discussion

Cette étude démontre l'efficacité d'une approche hybride pour la modélisation de la salinité de surface dans le canal du Mozambique. L'identification des précipitations comme forçage dominant par le modèle linéaire correspond aux mécanismes physiques connus dans cette région, où les apports fluviaux et les précipitations côtières influencent fortement le bilan d'eau douce. La performance du modèle neuronal ($R = 0,67$; $RMSE = 0,060$ PSU) met en lumière l'importance des interactions non linéaires entre les forçages climatiques pour une prévision précise.

Cependant, la capacité de généralisation du modèle au-delà de la période d'apprentissage et à d'autres sous-régions du canal nécessite des validations supplémentaires, notamment en intégrant des données à plus haute résolution spatiale et en considérant des processus locaux tels que la dynamique des courants ou les effets de la topographie côtière. Les projections obtenues (2019-2028), bien que cohérentes, restent dépendantes des scénarios de forçages utilisés et gagneraient à être confrontées aux sorties de modèles climatiques régionaux.

Conclusion

Cette recherche a permis de développer et de valider un modèle hybride original pour la prévision de la salinité de surface dans le canal du Mozambique. La combinaison d'une régression linéaire interprétable et d'un réseau de neurones performant a fourni un outil précis ($RMSE = 0,060$ PSU) et physiquement cohérent, capable de restituer la variabilité saisonnière et interannuelle de la salinité.

Les principaux résultats confirment le rôle prépondérant des précipitations et soulignent la contribution des processus non linéaires. Le modèle constitue ainsi une base solide pour des applications opérationnelles de surveillance régionale et ouvre des perspectives pour l'étude intégrée des interactions océan-atmosphère dans un contexte de variabilité climatique accrue.

Références

- [1] Bishop, 2006. « PATTERN RECOGNITION and MACHINE LEARNING »
- [2] Chatfield, C. (2004). *The Analysis of Time Series: An Introduction*. Chapman & Hall.
- [3] Coulibaly, P., Anctil, F., & Bobée, B. (2000). Daily reservoir inflow forecasting using artificial neural networks. *Journal of Hydrology*, 230, 244–257.
- [4] Cybenko, 1989 « Approximation par superpositions d'une fonction sigmoïde »
- [5] Emery, W. J., & Thomson, R. E. (2001). *Data Analysis Methods in Physical Oceanography*. Elsevier.
- [6] Glorot et al, 2011 « Réseaux neuronaux à redresseur clairsemé profond »
- [7] Hornik et al, 1989 “Multilayer feedforward networks are universal approximators”
- [8] R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE) ? *Geoscientific Model Development*, 7, 1247–1250.
- [9] Talley, L. D., Pickard, G. L., Emery, W. J., & Swift, J. H. (2011). *Descriptive Physical Oceanography* (6^e éd.). Academic Press.
- [10] Tomczak, M., & Godfrey, J. S. (2003). *Regional Oceanography: An Introduction*. Pergamon.
- [11] UNESCO (1981). *Background papers and supporting data on the Practical Salinity Scale 1978*.